DOCUMENT RESUME

ED 162 556 HE 010 554

AUTHOR McIntyre, Charles J.
TITLE Elevating the Importance of Teaching. National Project III. Final Report.
INSTITUTION Illinois Univ., Urbana. Office of Instructional Resources.
SPONS AGENCY Fund for the Improvement of Postsecondary Education (DHEW), Washington, D.C.
PUB DATE 31 Aug 77
GRANT G007502210
NOTE 17p.; For related document see ED 146 825

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *College Faculty; *College Instruction; *Effective Teaching; Evaluation Criteria; Evaluation Methods; *Faculty Evaluation; *Faculty Promotion; Higher Education; Information Systems; Instructional Improvement; *Peer Evaluation; Program Descriptions; Program Evaluation; Questionnaires; State Universities; Teacher Salaries
IDENTIFIERS Program Evaluation Survey; *University of Illinois

ABSTRACT

The University of Illinois' project on the elevation of the importance of teaching consisted of two programs: one to try to better systematize information about teaching effectiveness, and one to develop and validate a peer evaluation questionnaire to secure interpretable information. It is believed that the importance of teaching effectiveness will be elevated if faculty plays a larger role in pay and promotion decisions, and this is an important factor in the project. The development of the project is outlined, and one item is appended: Issue 4 of "Criteria for the Evaluation, Support, and Recognition of College Teachers," dealing with the evaluation of college teachers. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

FINAL REPORT

National Project III

ELEVATING THE IMPORTANCE OF TEACHING

Office of Instructional Resources
University of Illinois
Urbana-Champaign

August 31, 1977

Charles J. McIntyre
Project Director

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Submitted to:

Fund for the Improvement of Post-Secondary Education
Department of Health, Education, and Welfare

Grant Number: G007502210

Organization of the Report

The University of Illinois's project on the elevation of the importance of teaching actually consisted of two projects, both clearly related to the general objective, but so significantly different that it appears best to report them in two clearly different portions of this final report.

The first project had to do with trying to better systematize information about teaching effectiveness so that teaching could have more systematic weight in pay and promotion decisions. A description of the rationale, procedures, and present status of the work we have done on individual teaching evaluation constitutes the first part of this report.

Concurrently, this campus has been heavily concerned with evaluation, in all of its aspects, of <u>departmental</u> and <u>program</u> effectiveness. One source of information for these programatic evaluations—and one having particular relevance for elevating the importance of teaching—had to do with the systematic, organized, and interpretable input from students who were majoring in the departments being studied. The development and validation of a questionnaire—the Program Evaluation Survey—to secure information of this kind was our second project. This is dealt with in the second part of this report.

I. <u>Evaluation of Teaching Effectiveness</u>

It is our belief that the status of teaching will be elevated if the teaching effectiveness of faculty plays a larger role in pay and promotion decisions. This is an especially important consideration in an institution in which research and scholarship may appear to be more highly valued than teaching. A significant aspect of the problem is not so much the <u>intent</u>

to value teaching less highly, but rather the difficulty in documenting indices of teaching effectiveness. The importance of teaching, thus, often seems to be undervalued because of the lack of evidence upon which discrimations may be made.

As one source of evidence, the University now has various instruments which are used to elicit student evaluations of teachers. A new one is in development which, it is hoped, will more clearly discriminate between student evaluations that are to be used for pay and promotion purposes and those which are to be used for self-improvement.

It has been clear to us, however, that student evaluations, whatever may be their merits, need to be supplemented by some form of colleague evaluation. The problem is to work out a colleague evaluation system which is credible, practical, and seen to be fair. Without minimizing the value of student questionnaires, this project has concentrated upon developing criteria and methods for obtaining colleague evaluations.

A campus Ad Hoc Committee on Teaching and Service was appointed by the Vice Chancellor for Academic Affairs to advise the Vice Chancellor and the Campus Committee on Promotions and Tenure.

During the year preceding the first year of the FIPSE grant, the powerful Campus Committee on Promotions and Tenure formally pointed out to the Vice Chancellor that it was exceedingly difficult to give proper weight to excellence in teaching or in service in their deliberations and recommendations, because evidence of teaching and/or service excellence was usually presented in a rather disorganized, hapazard fashion and was thus

very hard to weigh and evaluate. The Campus Committee on Promotion and Tenure asked the Vice Chancellor to provide guidance to the campus as a whole in collecting and presenting information on these matters that could have the same kind of systematization and credibility which the information on scholarship, productivity, and creativity normally do.

The Vice Chancellor's response to this request was to appoint an Ad Hoc Committee on Teaching and Service[1] whose charge included the following.

> "The University is usually described as having three basic roles: teaching, research, and service. The campus Committee on Promotions and Tenure has noted repeatedly that the recommendations for promotion they examine often do not present well-documented records of teaching and service while the case for research or creative artistry is almost always exhaustively presented. Since the quality of a person's teaching and service should play a role in the deliberations concerning promotion and tenure, the fact that the evidence presented in many instances is sparse is troublesome to say the least.
>
> "I should make it clear that I am not asking the Committee to develop procedures or a rationale for a multi-track system in which one might be promoted on the basis of research or <u>teaching</u> or <u>public service</u>. In my view (and the campus Committee on Promotions and Tenure appears to agree) research, scholarship, and creative artistry of high quality must be demonstrated if one is to successfully be recommended for promotion or for indefinite tenure. However, in many instances outstanding performance in teaching and/or service, if the case is well documented, can result in a favorable outcome when the scholarship or creative artistry is itself not quite sufficient.

---

[1] It will be noted that the Committee was concerned both with teaching and service although the project being reported was concerned only with teaching. As it happened the Committee has concerned itself with these two aspects separately so that the present report will deal with the Committee's work on teaching only.

> "In summary, we need the guidance of knowledgeable individuals in exploring better ways of evaluating teaching and service. Hopefully, such guidance could be passed along to departmental executive officers to aid them in preparing promotion papers, or more generally in evaluating the progress of individuals."

The Committee had the following initial membership:

Dean, College of Applied Life Studies[1], Chairperson
Professor, Social Work
Professor, Psychology
Professor, Physics
Professor, Veterinary Medicine
Professor, Plant Pathology
Professor, Music
Professor, Business Administration
Graduate Student

The Committee also had as permanent resource persons four individuals from the Office of Instructional Resources who are professionally concerned with questions of this kind and an Associate Vice Chancellor for Academic Affairs responsible for Continuing Education.

The Committee has met regularly since November, 1975 spending much of its effort in reviewing problems and recommendations brought to it by individual members, who draw either upon their own experience or the procedures of the departments with which they were associated. The Committee also reviewed, criticized, debated, and edited documents prepared by one or more of its members.

In January of 1977, the Committee had ready a draft statement which it asked the Vice Chancellor for Academic Affairs to distribute to the campus

---

[1] During the course of the Committee's work, its chairperson was killed in a flying accident. The Director of the FIPSE project was appointed Acting Chairperson and is still serving in that capacity.

for comment. This he did by memorandum to Deans, Directors, and Department Heads, pointing out the reasons for the report, and the desirability for comment from as broad a base as possible. Some department heads took it upon themselves to distribute the report to every faculty member, most did not. Most, but not all, colleges had the report reviewed by a committee and in due course we received comments back from some committees, some deans, some department heads, and some individuals. We were frankly rather surprised at the sparseness of the response, especially considering the extent to which the report potentially effects the vital interests of every individual. Most of those responses which were received criticized certain aspects of the report. Some were broadly commendatory; some as broadly damning. It is the Committee's belief that most of those criticisms can be dealt with in the final draft.

The Committee's Chairman has spoken informally to a few deans and department heads who had not responded, surprisingly he thought, to determine the reason. In general, he was told that it seemed as though something along these lines was necessary and in one way or another was almost certain to be forthcoming eventually, and that what the Committee was proposing seemed to be generally acceptable. Hence, no response was necessary. The Chairman is concerned that another reason for non-response is the relative innocuousness of the report itself. That is to say, it may at this stage be too general to engage the interest of the faculty.

In any event, a final draft will be developed within the next few weeks and submitted to the Vice Chancellor for Academic Affairs for general distribution to the campus. Meanwhile, as it will be seen below, a very similar document has been prepared and distributed as Criteria VI, which is an immediate practical outcome of this phase of the grant.

While the Committee was doing its work, described above, there were several other concurrent activities occurring on the campus which were helpful.

A. Intensive interviews with a substantial sample of department heads were conducted to determine how they perceived instruction to be evaluated and how they thought it should be evaluated. The reports of these interviews are internal, confidential papers, but material drawn from them has been used to inform and guide the Committee in its work.

B. In the fall of 1976, one of the main topics of a conference held annually involving approximately 100 faculty of this University had to do with improvement of instruction. One of the outcomes of the conference was that the faculty pointed out that more was needed to be done to relate promotion and other rewards to teaching excellence in order to encourage more interest and effort to be spent on improvement.

C. At a conference on future development, also in 1976, the Executive Committee of the College of Liberal Arts and Sciences recommended, "That teaching performance be weighted more heavily in both promotion and salary increases.... We reaffirm we want to give more credit to good teaching. We realize that course evaluation questionnaires are not adequate in themselves.... We want our faculty to not only do quality scholarship but also inspired teaching."

D. Two committees appointed by the Board of Trustees on tenure matters reported in 1976. One, a faculty committee, seemed

to conclude that tenure policies as presently implemented are satisfactory. The other committee, a student committee, came down heavily on the side of teaching being undervalued in the reward structure. The Board of Trustees in a formal action reaffirmed its dedication to excellence in teaching and noted that more needed to be done on the validity and reliability of methods of assessing teaching effectiveness.

One negative experience which we had that others may find useful was as follows: We initially intended to form committees in numerous departments to work with us in trying to develop guidelines which seemed to be appropriate in their discipline. During the first year, committees were appointed in Educational Psychology, Music, Geography, and the School of Life Sciences. We had anticipated that there would be many more committees during the second year of the project. Unfortunately, we had to give up this mode of operation as impractical. Although there was no particular difficulty in having committees appointed, we found it so difficult as to be unfeasible to actually get these committees to meet and do serious work on the problem. While probably an over-simplification, our diagnosis of the situation is that senior members of the committees were not really very much interested in the problem posed to the committee. Those who were junior, although interested, doubted that their work on the committee would affect promotion policies in time to matter to them and that, with promotion so difficult to obtain here now, they were better off spending their time otherwise.

One final—and quite significant—development needs to be reported. That is that the work of the Committee on Teaching and Service of the University of Illinois served as the basis for a publication coming out of National Project III, entitled, Criteria VI, "Evaluation of College Teachers." Criteria VI was revised from the University of Illinois's draft report by the Project Director, with substantial assistance from Professors Seibert of Purdue and Owens of Kansas State University, and with review by the other representatives of the Project consortium.

In my judgment, that report is essentially the same as will be disseminated in the next month or two as the University of Illinois Committee's report to our own campus. The major difference will be that Criteria VI is more generalized and less specific to the University of Illinois. Hence, for general purposes, it is believed that Criteria VI will be more useful than the University of Illinois's report for institutions other than the University of Illinois. For this reason, Criteria VI is an attachment to this final report as one of the principal outcomes of this project. The report of the University of Illinois's Committee on Teaching and Service will be available later on request. That part dealing with teaching should be available in the fall, 1977, and the report on service and continuing education in the spring, 1978.

II. The Program Evaluation Survey

The second project which was undertaken had to do with the development of a program evaluation survey—an instrument to secure systematically student opinions about their major field.

The University of Illinois was engaged in a large-scale project involving the systematic and periodic evaluation of all academic departments. The Council on Program Evaluation, which was responsible for that activity, asked that a questionnaire be designed for the systematic collection of relevant data from student majors in the department being studied. The questionnaire—the Program Evaluation Survey—was to serve two functions: (a) University level administrators and the Council could use the information in making comparative judgments about many educational units, (b) faculty and departmental level administrators could use the information in identifying the relevant strength and weaknesses within each department so that program improvements could be implemented if needed. Three forms of the Program Evaluation Survey have been administered to fifty-three academic departments since 1974. In its current and (apparently) final version, it has twenty-four items plus four open-ended questions.

The instrument has been used for both formative and summative evaluation, with its contribution to formative evaluation being somewhat stronger than its contribution to summative evaluation. Several tests of validity have been made, i.e., structural validity, discriminant validity, and criterion-related validity. Each kind of analysis yields somewhat different results but together they indicate sufficient validity to argue for the usefulness of the instrument if used with care. Factor analysis yielded four factors: Satisfaction with mentorship, personal value, satisfaction with instruction, and satisfaction with curricular offerings.

A monograph entittled "The Development and Use of the Program Evaluation Survey (PES)" fully reports this work and is included as Attachment II.

CRITERIA FOR THE EVALUATION, SUPPORT, AND RECOGNITION OF COLLEGE TEACHERS

A SPECIAL PUBLICATION OF THE FUND ASSOCIATES IN NATIONAL PROJECT III, PREPARED AT THE CENTER FOR RESEARCH ON LEARNING AND TEACHING, THE UNIVERSITY OF MICHIGAN

Number VI

May, 1977

Evaluation of College Teachers

This paper will discuss the evaluation of teaching and teachers, primarily for the purpose of salary, promotion, and retention decisions. It will overlap somewhat with Criteria II which was concerned with evaluation by students, both for this purpose and for the related—but different—purpose of providing information to the professor for use in improving instruction. In considering student ratings here, we focus on the administrative purpose, i.e., to assist in pay, promotion, and retention decisions.[1]

Although student ratings can be important, they are not sufficient for administrative evaluation; therefore, we will also give considerable attention to evaluation of faculty by their peers.

Institutional reward structures differ significantly in the relative importance assigned to excellence in teaching and to excellence in research. Many institutions, especially the community colleges and the liberal arts colleges, give first weight to teaching, in fact as well as in what they profess. Others, especially those that profess equal concern for the great triumvirate of teaching, research, and service are often perceived by the faculty as, in fact, making various indices of scholarship and creativity the only ones that really matter. This is quite understandable on at least two counts: (1) Society's rewards for institutions, for departments, and for individuals come primarily through activities that can gain public recognition and evaluation, even if that "public" consists only of professional peers; (2) Institutions which have this means of evaluation available to them are more comfortable with it than with the more sensitive (and possibly more abrasive) evaluations that need to be made of the process of teaching. Furthermore, there is no evidence that institutions which reward primarily on the basis of research prominence provide better or worse instruction than those which reward primarily on the basis of instructional excellence. One might therefore erroneously conclude that success in research is highly related to success in teaching. It is not.

Nevertheless, our position is that institutions of every variety would be well served by adopting and maintaining a system for evaluating teaching proficiency. Such systems need to be developed so that the parties involved regard them as sensitive, impartial, and practical to maintain. Their goal should be to recognize directly the status of superior accomplishments in teaching and to reflect this recognition in official statements regarding its status.

Whatever the system, the first essential is to make it generally known and understood. Institutions should make as explicit as possible the criteria upon which administrative decisions will be made. Institutional styles of governance differ, but we recommend that the criteria be developed and agreed upon by discussion and negotiation among the parties concerned. These criteria should allow for individual differences within faculties to the extent that such differences are consistent with the unit and institutional mission and values. In general, too, it is best to develop criteria within the smallest practical unit. "The smallest practical unit" in some cases may be the whole institution, but frequently it will be a unit like a department or a group of departments perceiving themselves as being essentially similar. In any case, the unit ought to be explicitly defined and given responsibility for developing evaluation methods and criteria agreeable to its members. One reason for this is that different disciplines often require different methods and criteria for evaluation. Beyond this, the credibility and acceptability of such systems require that faculty members have a strong hand in their development. To insure compatibility within the system, standards adopted at each level will need to be reviewed, understood, and accepted by the next higher level.

There remains, of course, the problem of how to document excellence in teaching. Several generalizations seem relevant. One is that it is probably fruitless to search for easily applied criteria or for a single set of criteria that will be applicable to every case. This is unfortunate but need not be daunting. Teaching is such a significant part of our institutional mission that it deserves the effort that will have to go into recognizing it.

Secondly, since such evidence as can be gathered is likely to be partial and incomplete, each piece of evidence probably will need to be considered in relation to the whole. Thus, for example, student ratings need to be considered in light of peer opinions, characteristics of courses which were taught, and so on.

Thirdly, these proposals may make the job of promotion committees harder, although not really different, than it is now for those committees that normally pay little attention to teaching. Such promotion committees have always had to deal with information on scholarship, research, and creative artistry from evidence that is partial and incomplete; they have had to judge and weigh that evidence. These proposals im-

[1]Evaluation has at least two purposes which sometimes cannot be (and perhaps should not be) wholly separated. One purpose is for recognition and reward; the other is for personal growth and improvement. (In some situations a third purpose might be added—to provide information for "consumers," i.e., students.) If a faculty has the objectivity, mutual trust and self-confidence needed in order to use evaluation for both reward and growth in mutually supportive ways, this is ideal. However, it is probable that those purposes will more often be accomplished better if they are kept separate.

ply that faculty members and administrators making decisions about promotion and tenure should expect to put as much care and sensitivity into the interpretation of information about teaching as they do into considering other evidence of scholarship upon which such decisions are made.

While we deal here mainly with proposals for evaluating classroom and other direct instruction, distinguished contributions to teaching can also be found in such teaching-related activities as out-of-class instruction, instructional and curricular development, instructional research, and advising. Distinction in any of these areas may be the unusual case, but where it occurs it should be recognized, and poor performance in any of these activities should be weighted negatively.

Direct Instruction

Our discussion of direct instruction will be concerned mainly with **factors to be considered** in evaluating direct instruction and will not at this time give much attention to **how to obtain or systematize** that information. A great deal of work is needed in these matters, especially as it relates to information supplied by faculty peers. Faculties will need to consider not only the relevance of the criteria suggested but also how to collect and systematize the data in their situation.

The most commonly available systematic source of information is questionnaires administered to students about instruction and instructors. Less often used, but fully as important, other types of evaluative information will be discussed later under the categories of: (1) Instructor's own statement of his or her teaching objectives and significant teaching activities; (2) Instructional materials; (3) Evaluation and grading methods; and (4) Instructor attributes.

We believe that students are an important source of information about teaching and that questionnaires are a valuable means of eliciting some kinds of information. We recognize, however, that there are problems and dilemmas with respect to using student questionnaire responses in administrative decision-making. These are discussed below.

Student Ratings

First, some assumptions or generalizations about student ratings: (1) Students can generally judge reliably how satisfactory an instructor and course have been for them.[2] (2) Students are not qualified to judge matters requiring professional background, e.g., whether the material presented is current, whether it is placed at the proper level of difficulty, whether it is properly integrated with the curriculum as a whole, or whether the instructor's evaluation procedures are appropriate and technically adequate. (3) Standardized questionnaires are not useful in evaluating all kinds of instructors. In particular they have generally been of only limited value in studio courses, clinical practice, self-paced courses, team teaching, and individualized instruction.

Given the above assumptions, a number of problems in the use of student evaluations may be unsolvable if one requires near perfect solutions. However, the dilemmas are far from insurmountable when questionnaire results are used with care and understanding. In fact, we observe something of a contradiction in the way faculty members consider different kinds of information used for evaluative purposes. They tend to accept the usual evidence of scholarship, such as publications, even though very little is known about the implied validity, reliability, freedom from bias, and comparability of such evidence. However, many are highly critical of similar "softness" in student ratings and may for this reason discount them altogether.

The better posture is to understand the limitations inherent in student ratings and then subject that data, also, to the interpretive and judgmental process which is customarily applied to the more usual evidence of scholarship.

Some of the most persistent questions and doubts about the usefulness of student questionnaires seem to center on the following issues.

The Different Purposes for Student Evaluation Questionnaires

Student evaluations are used for two main purposes—diagnostic (i.e., for self-improvement) and administrative (for pay and promotion considerations).[3] These purposes are not necessarily compatible. (Much of the problem with course evaluation questionnaires has stemmed from expecting them to serve both purposes well.) The information that faculty members need in order to help improve their instruction usually is specific to their courses, their intentions, and their students. There is no reason to suppose that what one instructor wants to know or will find useful will be the same for other instructors. Conceivably, for this "formative" (improvement) purpose, a unique set of questions could be prepared for every course. Several systems now exist which can provide such flexibility. Typically, they are computer-aided and they allow instructors to select perhaps 30 or 40 rating items from a large and varied item "pool" and, in effect, to create tailor-made rating instruments. Such systems are now in use or under development at several dozen institutions, including Purdue, The University of Michigan, University of Wisconsin-Whitewater, University of Alabama at Birmingham, and the University of Illinois.

For administrative purposes, "summative" evaluation is more appropriate. In this instance, "common" measures are needed since it is necessary to compare faculty members. Some highly generalized questions are most useful since they tend to be equally applicable across a wide range of courses. Such questions can be quite simple, such as:

Rate the course content.
Rate the instructor's overall teaching effectiveness.
Rate the overall value of the course to you.
Rate the amount you have learned in this course.

Questions such as these tend to be highly reliable[4] and we recommend that this type appear on every questionnaire which is intended for administrative use. Such questions are not very helpful to the instructor

[2]For example, analysis of University of Illinois Course Evaluation Questionnaire data shows that the majority of CEQ **subscale** reliabilities for class sections are greater than .65 which is an acceptable level of reliability. The CEQ **total** score yields a much larger proportion of class section reliabilities, about .65. Similar data are reported by other institutions for similar questionnaires.

[3]As mentioned previously, there is frequently a third purpose—namely to provide information to prospective students.

[4]Reliability in this case refers to class sections mean reliability over a large number of classes. The general range of reliability for each of these items is from .80 to .94. Other investigators report similar reliabilities for similar items.

seeking self-improvement hints, but that is not their purpose. Some institutions already have normative data on thousands of sections for these kinds of general questions.

Aside from the fact that the two purposes require different kinds of questions, their basic incompatibility is suggested by the personal meaning they have for the faculty member. Under "improvement" conditions, participation is usually voluntary, feedback is confidential, and the participant's motivation is to become more effective. Administrative use is much more threatening: participation is usually required, feedback goes to the individual's superior, and such crucial matters as salary, retention, or promotion will be affected by the results. Only the rare faculty member can view the same instrument as a "friendly advisor" and as a determiner of his or her professional future.

Validity

It is reasonable to ask, "If the questionnaire purports to measure teaching effectiveness, does it?" The question is deceptively simple. It implies that some measurable, independent, generally agreed upon outside criteria exist where in fact they do not. It is quite impossible to get academics to agree on a general formulation of what constitutes teaching effectiveness, and perhaps there is no reason why they should. What is applicable in one case need not be at all applicable in another.

Instead, we suggest that the results of student evaluation questionnaires be looked upon as answering a somewhat different question, namely, "To what extent do students **think** that this instruction was effective?" We believe that this is a meaningful and important question in considering faculty merit and that student evaluation questionnaires can provide a reliable answer to it. The general feelings of satisfaction and success which students gain from their courses constitute one dimension of effective teaching. While high student ratings do not guarantee that the teacher is effective, it will be a rare teacher who is truly effective and still receives low student ratings.

Norms

Absolute scores, such as a number or a description on a scale, are difficult to interpret unless one knows how other people are distributed on the same scale. Here we are faced with a dilemma: Who are the others? What is the norm base? On most campuses, norms have generally been developed from faculty members who have elected to have their courses evaluated and, therefore, those norms may not be representative of the faculty as a whole.

In addition, there are problems of norming which seek to take into account variables such as class size, whether the course was required or elective, grade distribution, and level. However, general items such as those on page two tend to be less sensitive to these kinds of variations than are more specific items.

To overcome the more common weaknesses of existing norms, it is frequently suggested that student evaluations be made mandatory in all courses for all instructors. But before this seemingly simple remedy is adopted, its full implications, and alternatives to it, should be exained. First, for many institutions, full faculty cooperation with such mandatory schemes is practically impossible. At times, cooperation may appear to be given when in fact only perfunctory implementation is achieved. Thus, the quality of the norm base may not improve, even though the accumulated numbers become larger.

Three additional considerations may recommend against full, mandatory use of student valuations: cost, faculty morale or rapport, and student tolerance. Costs can be unexpectedly high since, conservatively, each student's evaluation of a course will cost the institution 15¢ and, more realistically, it will cost twice that amount—or even more. Thus, multiplied over all students in all courses during each semester or term, annual costs can reach well into the thousands of dollars. Do the benefits justify such expense? Faculty morale and rapport with faculty can also suffer—especially if faculty are required to use an evaluation form that is common to all courses and, as a result, not very sensitive to the characteristics of some. ("Why do they persist in asking about the quality of my lab teaching when there is no lab?" or "Why don't they ask about the required readings, which are the heart of the course, rather than the textbook, which I know is inadequate?") In addition, student tolerance may be exceeded if each student must evaluate courses and instructors four to six times each semester. Although hard evidence is lacking on this, it is at least plausible that students who must repeatedly evaluate courses will lose interest in giving careful attention to the task. Obviously, for evaluations to be useful, they should represent judgments reached with much thought and care, so conditions conducive to this need to be maintained.

Thus, there are several potentially serious problems associated with total and mandatory use of student course evaluation procedures. Consequently, it is recommended that institutions develop some sampling procedure which is feasible under local circumstances and which strives to include all appropriate courses and instructors from time to time. Voluntary participation for whatever reason should, of course, be encouraged, but beyond that the plan should seek coverage of courses and instructors which is broadly representative but which remains sensitive to matters of data quality, cost, faculty morale, and student tolerance.

Other Kinds and Sources of Information

We have said that student ratings are an important but not sufficient source of information about instruction. One needs corollary data, in part to provide information which students are unlikely to have or to be able to make good judgments about, and in part to provide checks of internal consistency or explanations for perceived inconsistency. This kind of information, usually obtained from one's peers, is as important and requires collection procedures as systematic as for evaluations from the students.[5] As a general framework, we suggest direct instruction be evaluated on the basis of the following categories:

Instructor's own statement of teaching objectives and significant teaching activities
Instructional materials
Evaluation and grading methods
Instructor attributes

[5]It is perhaps unfortunate that relatively little work has been reported on the systematic collection of information from former students—those who have graduated and can make judgments from the perspective of greater experiences and maturity than current students.

These four categories are not mutually exclusive or exhaustive, and certainly not all would necessarily be present in any given classroom setting. The outline of subcategories contains questions which might be asked in the evaluation of instruction.

Instructor's Own Statement of Teaching Objectives and Significant Teaching Activities

This category asks the instructor to reflect on significant teaching activities. The intent is to explore a dimension often overlooked: that is, of what in the teaching performance is the instructor most proud—and is this consistent with the data collected from other sources? Data in this category might consist of the instructor's written response to brief, open-ended, written inquiry. It provides the instructor with the opportunity to explain strategies, objectives, or circumstances which made his or her effort unique. Was a particularly difficult objective being emphasized? Were facilities and equipment adequate? Were new approaches explored? Was there an important outcome which would not be observed through other appraisal procedures?

Is the statement consistent with other collected data? Is it reflective of weaknesses that appear in other forms of evaluation? Does it offer suggestions for self-improvement pegged to an appropriate future time schedule? Does it offer significant reasons to explain contradictory information obtained elsewhere?

Instructional Materials

The concern is with characteristics that describe the content of the course. Some of the characteristics to be judged and some guiding questions are given below.

- Textbooks, Handouts, and Syllabi
  Are they current, relevant, suitable to the course outline? Have new or creative teaching aids been developed or explored? Are they of a difficulty and quality appropriate to the course level?
- Lecture Notes
  Do they follow the course outline? Are they well organized, individually and as a sequence? Do they supplement (or needlessly duplicate) the text and other class materials?
- Reading and Reference Lists
  Do these supplement the lecture notes and class discussion? Are they up-to-date? Do they "challenge" the student?
- Non-reading Assignments
  Are sample problems and laboratory reports adequately presented to broaden the students? Are the projects coordinated with the rest of the syllabus material?
- Statements of Objectives
  Do these fit in with the overall departmental objectives? Has the instructor been able and willing to extend the usefulness of the course by integrating it with others which may have come before or will come later?

Evaluation and Grading Methods

Even though evaluation procedures interact with both instructional materials and instructor attributes, it is possible to make inferences about teaching quality by reviewing the formal evaluation process. Under this category we include: prompt and useful feedback from tests, clarity of test objectives, relationship of exams to objectives and other instructional materials, reliability and validity of tests or other measures of student achievement, fairness of grades, and rationale for grade distribution and for grading policy. Although the judgment of colleagues is of most importance, student comments should be obtained also.

- Examinations, Tests, and Quizzes
  Do they appear to be fair? Is the grading objective? Are thought questions blended with tests of memory? Are tests returned promptly and wrong answers explained? Do the grades follow a distribution pattern that is acceptable? Is there evidence that most students are achieving objectives?
- Student Projects
  Are they a relevant learning experience or simply "busy" work? Do they fit within the course objectives? Do they help solve a problem? Do they better prepare the student for life beyond the classroom?
- Grading
  On what standards is it based? Do the students know the standards? Is the grading fair?

Instructor Attributes[6]

The concern here is with less tangible aspects of in-class instruction and includes, but is not limited to, the following:

- Instructor Knowledge
  Although recognized as a prerequisite to effective instruction, substantive expertise is not necessarily a permanent attribute. Evidence that the faculty member has conscientiously kept up in the discipline and extended his or her own skills and understanding is thus relevant to teaching effectiveness. Likewise, if the instructor is responsible for instruction in areas outside his or her specialty, evidence that the necessary expertise has been acquired should be sought.
- Method of Instruction
  This refers to the preparation of the instructor for class and the development of systematic procedures for facilitating learning. Specific characteristics may include the following: presents material clearly, stimulates questions in class, uses class time efficiently, presents relevant examples, stresses important materials, changes methods to meet new situations, helps students gain new viewpoints and appreciation, leaves students feeling that class effort is worthwhile, and increases students' skills in thinking. In obtaining evaluative information on this topic, it is reasonable to assume that colleagues and administrators may be adequate judges of appropriateness of methods but students may make useful judgments on the effectiveness (outcomes) of different methods.
- Concern and Enthusiasm for Teaching
  The instructor's enthusiasm for teaching includes willingness to teach and to take a personal interest in the improvement of his or her own teaching. Specific concerns include commitment to the field of study, efforts to improve course quality, analysis of previous experiences, experimentation with new ways of teaching, search to improve quality of student achievement, demonstration of interest in material.

[6]It will be evident that one way to observe the behaviors described in this section is by visiting the classrooms of those being evaluated. In fact, this appears to be a relatively common procedure, especially on the part of administrative officers who use it as a primary evaluation method. Classroom visitation can be useful but it must be recognized that it is accompanied by severe problems of sampling, rater bias, and rater unreliability.

suggesting appropriate readings, relating course content to recent developments in the field, and interacting with colleagues on matters related to teaching and its improvement. Both colleagues and students can appropriately share in the evaluation of this area. Judgments of effectiveness should be tempered by consideration of the number of classes and students met, the number of different courses taught (and therefore preparation time and breadth required), and the difficulty of effectively teaching the courses assigned.

- Relations With Students
  The concern here is for instructor interaction with individual students. The instructor should strive to establish relationships with students that will promote student learning, maintain congenial relationships with students, and sensitize the teacher to student needs. Displays of courtesy and interest are also relevant; e.g., keeping appointments, relating positively to students, giving consideration to student suggestions, helping students having difficulty, and being fair and impartial in dealing with students. For evaluative information on this aspect, students probably should have the major input, though colleagues can often observe relevant behavior.
- Peer Evaluation of Scholarly Quality
  How do the individual's peers evaluate his or her knowledge, communication skills, and enthusiasm on the basis of performance in colloquia or seminars which the peers attend?

Out-of-Class Instruction

Academic Advising

Here we are concerned primarily with thesis guidance, supervision of individualized study and supervision of graduate students in research projects. Is the faculty member an effective resource person and mentor? How do the person's peers evaluate the quality of his or her students' work? Their defense of their work?

Instructional and Curricular Development

It is desirable to recognize leadership in formulating and developing a curriculum (usually a committee activity) or a course. Contributions meriting a positive evaluation include a superior reconceptualization of content, a beyond-the-ordinary marriage of content to method, or the development of an unusually successful textbook. Too often the creativity and effort which go into the development of new and better ways of teaching go relatively unnoticed and unrewarded. Innovative teaching and instructional strategies which produce demonstrably superior results should be as obviously encouraged as other noteworthy professional accomplishments. See **Criteria V** for a more detailed account of these contributions.

Instructional Research

Some faculty members conduct systematic studies related to the teaching of the subject matter in their discipline. Some departments would weigh published research of this kind heavily under the general rubric of research and scholarship, but others would not. Those departments might be more comfortable with considering such research not as a contribution to the discipline but as a contribution to teaching.

Personal Advising

There are two major qualities of the effective advisor: "interest" in students and "competence" as an advisor. The former can be inferred by observing such things as courtesy and warmth in interacting with students, the frequency with which advisees seek assistance on other than routine matters, conscientiousness in keeping appointments, or ability to call students by their first names (or to remember special things about them—hometown, hobbies, etc.)

Competence can be inferred by avoidance of enrollment errors; knowledge of rules, regulations, and procedures; soundness of advice regarding options and opportunities; use and appropriateness of referral sources.

Frequently, colleagues can judge some of these matters. Routine surveys of advisees (majors) can also provide helpful input.

Summary

We have reiterated not only the importance of teaching in higher education—an importance which most doubtless take for granted—but the importance also of explicitly recognizing teacher competence within an institution's formal reward structure. This is not easy to do, but we believe it can be done if faculty colleagues approach the task with the same sensitivity and the same recognition of complexity and richness which is generally involved in the valuation of scholarly credentials. An analogy might be to liken the promotion committee to a group of skilled musicians. They are already good at reading that part of the score pertaining to scholarship; they need to develop and bring that same expertise to interpreting the credentials which can be brought to bear on teaching.

Bibliography

Bergquist, W. H. and Phillips, S. R. **A handbook for faculty development.** Washington, D.C.: The Council for the Advancement of Small Colleges, 1975.

Centra, J. A. **Strategies for improving college teaching.** Prepared by ERIC Clearinghouse. Washington, D.C.: American Association for Higher Education, December 1972.

Doyle, K. O., Jr. **Student evaluation of instruction.** Lexington, MA: Lexington Books, 1975.

Eble, K. E. **The recognition and evaluation of teaching.** Washington, D.C.: American Association of University Professors, 1971.

French-Lazovik, Grace. **Evaluation of college teaching guidelines for summative and formative procedures.** Washington, D.C.: Association of American Colleges, undated.

Gage, N. L. "Student ratings of college teaching: Their justification and proper use." In Allen, Melnik, and Peele (Eds.)., **Reform renewal reward.** Amherst, MA: Clinic to Improve University Teaching, University of Massachusetts, 1975. Pp. 121-135.

Kulik, J. A. and McKeachie, W. J. "The evaluation of teachers in higher education." In F. N. Kerlinger (Ed.)., **Review of Research in Education,** Vol. 3. Itaska, IL: F. E. Peacock Publishers, Inc., 1975.

Miller, R. I. **Evaluating faculty performance.** San Francisco: Jossey-Bass Publishers, 1972.
Miller, R. I. **Developing programs for faculty evaluation.** San Francisco: Jossey-Bass Publishers, 1973.
Pace, C. R. **Evaluating learning and teaching.** San Francisco: Jossey-Bass Publishers, 1973.
Seldin, P. **How colleges evaluate professors.** New York: Blythe-Pennington, 1975.

This report was written by Charles J. McIntyre with assistance from Warren Seibert and Richard Owens. Based on a report by the Committee on Teaching and Service of the University of Illinois, Urbana-Champaign. SCE.

Activity which is the subject of this report was supported in whole or in part by the Fund for the Improvement of Postsecondary Education, Department of Health, Education and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the Fund and no official endorsement by that agency should be inferred.

For information about back file copies or multiple copies of the **Criteria** series, please write to Janet Lawrence, Assistant Director, National Project III, Center for Research on Learning and Teaching, The University of Michigan, 109 E. Madison, Ann Arbor, MI 48109

For further information about local activities in National Project III, write to:

Ball State University, 2000 University Ave., IN 47306—Project Director: Jerry J. Nisbet, Coordinator of University Evaluations

Bucknell University, Lewisburg, PA 17837—Project Director: Wendell I. Smith, Provost

Burlington County College, Pemberton-Brown Mills Road, Pemberton, NJ 08068—Project Director: Lee L. Schroeder, Director of Measurement and Evaluation

Earlham College, National Road West, Richmond, IN 47374—Project Director: Gerald R. Bakker, Department of Chemistry

University of Illinois, 205 S. Goodwin, Urbana, IL 61801—Project Director: Charles J. McIntyre, Director, Office of Instructional Resources

The University of Michigan, 109 E. Madison, Ann Arbor, MI 48109—Project Director: Stanford C. Ericksen, Center for Research on Learning and Teaching

Ohio Wesleyan University, Delaware, OH 43015—Project Director: Stephen C. Scholl, Dean of Educational Services

Purdue University, ENAD, Room 402, West Lafayette, IN 47907—Project Director: Warren Seibert, Director, Office of Instructional Services, Measurement and Research Center

State University of New York Central Administration, 99 Washington Ave., Albany, NY 12246—Project Director: Charles B. Neff, Assistant Chancellor for Special Projects

State University of New York, Oswego, Oswego, NY 13216—Project Director: David J. King, Dean for Evaluation and Research

Center for Research on Learning and Teaching
FIPSE Project
University of Michigan
Ann Arbor, MI 48109

Non-Profit
Organization
US Postage
PAID
Ann Arbor, MI
Permit No. 144

Designed by
University Publications
The University of Michigan